

Илья Дуванов

Структура базы данных

Обновлено 22 мая 2025, 16:51

Содержание страницы

- Структура базы данных
 - Общая информация
 - Описание таблиц
 - 1. users
 - 2. predictions
 - 3. actual_rentals
 - 4. models
 - Связь между таблицами
 - ERD-диаграмма

Структура базы данных

Общая информация

База данных проекта "Прогнозирование спроса на аренду велосипедов" разработана для хранения данных о прогнозах, фактических значениях аренд, моделей машинного обучения и пользователей.

Используется PostgreSQL как основное хранилище, обеспечивающее поддержку бинарных данных (`BYTEA`) для моделей и транзакционную целостность.

Структура базы данных поддерживает микросервисную архитектуру с разделением данных между различными сервисами (прогнозирование, обновление данных, дообучение моделей).

Описание таблиц

1. users

Таблица хранит информацию о пользователях системы, включая их идентификаторы и тип подписки.

Название поля	Описание	Тип данных	Ограничения	Примечания
user_id	Уникальный идентификатор пользователя.	UUID	PRIMARY KEY	Генерируется автоматически.
subscription	Тип подписки (free/premium).	VARCHAR(10)	NOT NULL	По умолчанию "free".
created_at	Дата и время регистрации.	TIMESTAMP	NOT NULL, DEFAULT NOW()	Отслеживание активности.

2. predictions

Таблица содержит записи прогнозов спроса на аренду велосипедов.

Название поля	Описание	Тип данных	Ограничения	Примечания
prediction_id	Уникальный идентификатор прогноза.	UUID	PRIMARY KEY	Генерируется автоматически.
user_id	Идентификатор пользователя.	UUID	FOREIGN KEY (users)	Ссылка на таблицу users.
request_id	Идентификатор запроса.	UUID	NOT NULL	Связь с клиентским запросом.
predicted_value	Прогнозируемое количество аренд.	INTEGER	NOT NULL	Результат модели.
timestamp	Время создания прогноза.	TIMESTAMP	NOT NULL, DEFAULT NOW()	Для анализа временных рядов.

3. actual_rentals

Таблица хранит фактические данные о количестве аренд, предоставленные пользователями.

Название поля	Описание	Тип данных	Ограничения	Примечания
actual_id	Уникальный идентификатор записи.	UUID	PRIMARY KEY	Генерируется автоматически.
request_id	Идентификатор запроса.	UUID	NOT NULL, UNIQUE	Связь с прогнозом.
actual_value	Фактическое количество аренд.	INTEGER	NOT NULL	Для валидации прогнозов.
timestamp	Время предоставления данных.	TIMESTAMP	NOT NULL, DEFAULT NOW()	Для анализа точности.

4. models

Таблица содержит сериализованные модели машинного обучения (общая и индивидуальные).

Название поля	Описание	Тип данных	Ограничения	Примечания
model_id	Уникальный идентификатор модели.	SERIAL	PRIMARY KEY	Автоинкрементный идентификатор.
user_id	Идентификатор пользователя (NULL для общей модели).	UUID	FOREIGN KEY (users) NULL	NULL для общей модели.
model_data	Сериализованная модель .pkl.	BYTEA	NOT NULL	Хранит модель в бинарном виде.
last_updated	Дата и время последнего обновления.	TIMESTAMP	NOT NULL, DEFAULT NOW()	Для отслеживания актуальности.

Связь между таблицами

- **users** → **predictions**: Связь один-ко-многим. Один пользователь может иметь множество прогнозов.
- **users** → **models**: Связь один-ко-многим. Один пользователь (или NULL для общей модели) имеет одну индивидуальную модель.
- **predictions** ↔ **actual_rentals**: Связь один-к-одному через `request_id`, так как каждому прогнозу соответствует одно фактическое значение.

ERD-диаграмма

